

# Implementing Data Standards

A brief overview

Jessica Utrup

Georeferencing for Paleo Workshop

April 28-29, 2020

**Yale** PEABODY MUSEUM OF NATURAL HISTORY

# Why do we need data standards?

- Institutions around the country and the world are working to digitize their collections.
- Different collections use different software and methods to capture their data.
- To gather this data and make it broadly searchable and shareable, we need data standards.
- Data standards allow data from multiple institutions and databases to be gathered and shared in a meaningful and coherent way.

# Within your institution

- Are there multiple collections within your institution?
- Do they all use same software?
- Do you share geographic fields?
- Are there controlled vocabularies for higher level geographic terms?
- Is georeferencing data shared for objects or specimens from the same geographic place?

# Controlled Vocabularies

- Many fields are easy to standardize.
  - Continents
  - Countries
  - States/Provinces
  - Oceans
  - Geologic Periods
    - The International Stratigraphic Commission (<http://stratigraphy.org/>).
  - Stratigraphic Groups
    - Many countries have an organization in charge of maintaining a standard lexicon.
  - Stratigraphic Formations
    - See Resources slide at the end of the talk for a small list to point you towards useful national geologic lexicons.

# Ways to Control Vocabularies

- Prepopulated lookup lists.
  - Only select folks have the ability to add to these lists.
  - Or everyone can add to it for a period of time and then the list is 'closed' to changes by most users.
- Best practices within your institution.
  - Will "County" in a county name be spelled out or abbreviated "Co."?
  - Will "Formation" in the geologic unit be spelled out or abbreviated "Fm."?
  - Will "Group" in the geologic unit be spelled out or abbreviated "Grp."?
  - A document with these types of guidelines should be discussed by a broad group within your institution and circulated to everyone.

# Making your data valuable

- The best way to make your data (and consequently your institution) more important is by standardizing your data to globally accepted standards.
- By utilizing an accepted standard your data becomes much more easily digestible by the big aggregators.
- Darwin Core is a set of widely used standards for both paleontological and neontological collections.
- By mapping your databases fields to their corresponding Darwin Core fields, you add value to your data!

# Darwin Core

Darwin Core is a standard maintained by the Darwin Core maintenance group. It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

# There's a Darwin Core field for that!

← → ↻ 🏠 🔒 dwc.tdwg.org/terms/#location ☆ a 🌐 🌍

## Location

- locationID
- higherGeographyID
- higherGeography
- continent
- waterBody
- islandGroup
- island
- country
- countryCode
- stateProvince
- county
- municipality
- locality
- verbatimLocality
- minimumElevationInMeters
- maximumElevationInMeters
- verbatimElevation
- minimumDepthInMeters
- maximumDepthInMeters
- verbatimDepth
- minimumDistanceAboveSurfaceInMeters
- maximumDistanceAboveSurfaceInMeters
- locationAccordingTo
- locationRemarks
- decimalLatitude
- decimalLongitude
- geodeticDatum
- coordinateUncertaintyInMeters
- coordinatePrecision
- pointRadiusSpatialFit
- verbatimCoordinates
- verbatimLatitude
- verbatimLongitude
- verbatimCoordinateSystem
- verbatimSRS
- footprintWKT
- footprintSRS
- footprintSpatialFit
- georeferencedBy
- georeferencedDate
- georeferenceProtocol
- georeferenceSources
- georeferenceVerificationStatus
- georeferenceRemarks

- Record-level
- Occurrence
- Organism
- MaterialSample
- Event
- Location**
- GeologicalContext
- Identification
- Taxon
- MeasurementOrFact
- ResourceRelationship
- UseWithIRI

---

- LivingSpecimen
- PreservedSpecimen
- FossilSpecimen
- HumanObservation

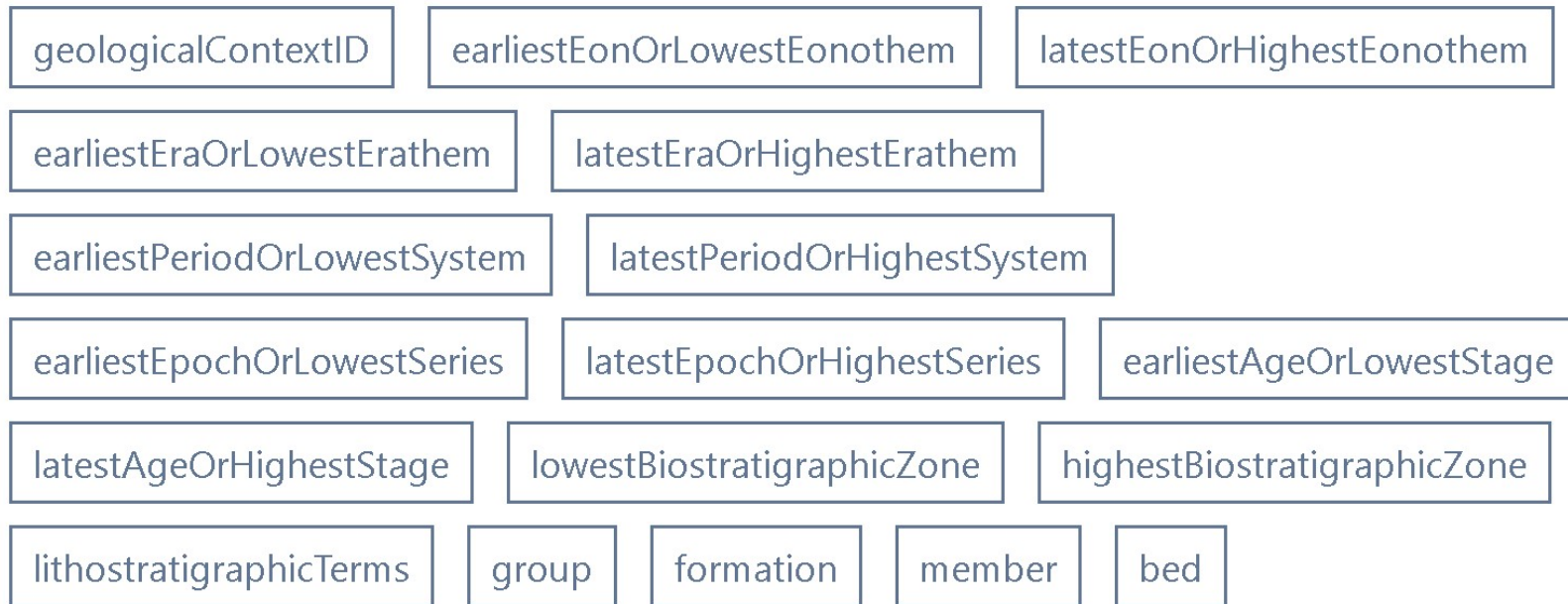


# There's a Darwin Core Field for that too!

← → ↻ 🏠 🔒 dwc.tdwg.org/terms/#geologicalcontext



## GeologicalContext



- Record-level
- Occurrence
- Organism
- MaterialSample
- Event
- Location
- GeologicalContext**
- Identification
- Terms

# Some Resources

- Darwin Core: <https://dwc.tdwg.org/>
- Getty Thesaurus of Geographic Names:  
<http://www.getty.edu/research/tools/vocabularies/tgn/index.html>
- Some Geologic Lexicons:
  - USA: <https://ngmdb.usgs.gov/Geolex/search>
  - UK: <http://www.bgs.ac.uk/lexicon/home.cfm>
  - Australia: <https://asud.ga.gov.au/>
  - Canada: [http://weblex.rncan.gc.ca/weblexnet4/Weblex\\_e.aspx](http://weblex.rncan.gc.ca/weblexnet4/Weblex_e.aspx)
  - Germany: <https://litholex.bgr.de/>
  - Mexico: [https://www.sgm.gob.mx/Lexico\\_Es/](https://www.sgm.gob.mx/Lexico_Es/)